THE JOINT DENSITY OF THE CARTESIAN COORDINATES OF NEAREST NEIGHBOURS UNDER A POISSON PROCESS IN N DIMENSIONS

O. Brian Allen and D. S. Robson

BU-633-M November 1977

ABSTRACT

Suppose that elements are distributed in n-dimensional Euclidean space according to a homogeneous Poisson process with density λ. Impose a Cartesian coordinate system on this space without reference to the actual realization of the process. Then it is shown that the joint density of the Cartesian coordinates of the element closest to the origin (that is, the nearest neighbour) is

$$\lambda e^{-\lambda \cdot V_n \cdot (x_1^2 + \cdots + x_1^2)^{n/2}}, \quad -\infty < x_j < \infty \quad (j = 1, \cdots, n)$$

when V_n is the volume of the unit ball in n-dimensions. It is noted that while the coordinates are uncorrelated for all dimensions $n \geq 2$, only in the case $n = 2$ are they independent. In this case, they have a normal distribution.

The joint density of coordinates of the k-th nearest neighbour is also computed.

The results are obtained by using the polar coordinate transformation in n-dimensions. This transformation, although non-linear, is tractable and very useful in dealing with problems involving spherical symmetry.

THE JOINT DENSITY OF THE CARTESIAN COORDINATES OF NEAREST NEIGHBOURS UNDER A POISSON PROCESS IN N DIMENSION

O. Brian Allen and D. S. Robson

1. Introduction and Summary

Suppose that individuals are distributed in n-dimensional Euclidean space according to a homogeneous Poisson process with density λ . Recall that the Poisson process is characterized by the following:

i) The number of individuals occurring in a region with volume v has a Poisson distribution with mean $\lambda \cdot v$.

ii) The number of individuals in non-overlapping regions is independently distributed.

We will superimpose a coordinate system on this process. We wish to do this without reference to the actual realization of the process (that is, independently of the actual realization), although the origin is sometimes chosen to be a "randomly" selected individual.

The joint density of the Cartesian coordinates $(X_1, \cdots, X_n)$ of the individual nearest the origin (the nearest neighbour or n.n.) is shown to be

$$\lambda e^{-\lambda \cdot V_n \cdot (x_1^2 + \cdots + x_1^2)^{n/2}} , \quad -\infty < x_j < \infty \quad (j = 1, \cdots, n)$$

where V_n is the volume of the unit ball in n-dimensions. It is noted that while the coordinates are uncorrelated for all dimensions $n \geq 2$, they are independent only in the case $n = 2$ where (X_1, X_2) is normally distributed.

The above results can be extended for the k-th nearest neighbour. The density of the Cartesian coordinates of the k-th nearest neighbour is

$$f^{(k)}_{x_1,\cdots,x_n}(x_1,\cdots,x_n) = \frac{\lambda^k(V_n)^{k-1}}{(k-1)!}(x_1^2+\cdots+x_n^2)^{\frac{n(k-1)}{2}} e^{-\lambda\cdot V_n\cdot(x_1^2+\cdots+x_n^2)^{\frac{n}{2}}}$$

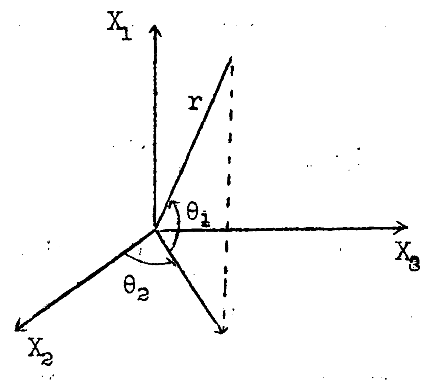
$$-\infty < x_i < \infty \quad (i=1,\cdots,n) \quad .$$

2. Polar Coordinates in n-Dimensions

In the derivations to follow, it will be convenient to make use of the polar coordinate representation of a point in n-dimensional Euclidean space. This is a generalization of the familiar polar coordinate representation of a point in the plane. The n-dimensional polar coordinates of a point are $(R,\Theta_1,\cdots,\Theta_{n-1})$ where $R \geq 0$, $-\frac{\pi}{2} \leq \Theta_j \leq \frac{\pi}{2}$ for $j = 1,\cdots, n-2$ and $0 \leq \Theta_{n-1} < 2\pi$. For notational convenience, set $\cos\Theta_j = C_j$ and $\sin\Theta_j = S_j$. The relation between the Cartesian coordinates of a point $(X_1,\cdots,X_n)$ and its polar coordinates $(R,\Theta_1,\cdots,\Theta_{n-1})$ is given by the following non-linear transformation.

$$\begin{aligned}
X_1 &= RC_1C_2 \cdots C_{n-2}C_{n-1} \\
X_2 &= RC_1C_2 \cdots C_{n-2}S_{n-1} \\
&\vdots \\
X_j &= RC_1C_2 \cdots C_{n-j}S_{n-j+1} \\
&\vdots \\
X_n &= RS_1
\end{aligned} \quad . \qquad (1)$$

The polar coordinates have the following geometric interpretation. R represents the distance from the origin to the point $\underline{X} = (X_1,\cdots,X_n)$. Θ_1 is the angle between $\underline{X}$ and its projection onto the subspace $X_n = 0$. Θ_2 is the angle between this projection and the projection of $\underline{X}$ onto the subspace $X_n = 0$, $X_{n-1} = 0$. This process of defining Θ_j by projecting $\underline{X}$ onto successively lower dimensional subspaces continues until $\underline{X}$ is projected onto the (X_1,X_2)-plane. Θ_{n-1} is the angle that this projection makes with the positive X_1 axis. An example in three dimensions is depicted below:



We will record certain facts without proof. For a more complete discussion see M. G. Kendall [1]. The Jacobian of the above transformation is

$$R^{n-1}C^{n-2} \cdots C_{n-2} \quad . \tag{2}$$

The volume of an n-dimensional ball of radius r is $V_n(r) = 2r^n \cdot \pi^{n/2}/n \cdot \Gamma(n/2)$. We will denote the volume of the unit ball by V_n . The surface area of an n-dimensional sphere of radius r is

$$S_n(r) = \frac{d}{dr}V_n(r) = 2r^{n-1}\pi^{n/2}/\Gamma(n/2) \ .$$

We shall denote the surface area of the unit sphere in n dimensions by S_n .

3. Development of Joint Density of $X_1, \cdots, X_n$

Our approach will be to derive the density of the polar coordinates of the nearest neighbour (n.n.) and then to transform back to Cartesian coordinates. First, we find the marginal density of R.

$$P\{R > r\} = P\{\text{no individuals in a ball of radius } r\}$$
$$= e^{-\lambda V_n r^n}$$

by the first property of a Poisson process. Hence the density of R is

$$f_R(r) = \lambda \cdot V_n \cdot n \cdot r^{n-1} e^{-\lambda \cdot V_n r^n}, \quad r \geq 0 \quad .$$

To obtain the density of the remaining polar coordinates, namely $\Theta_1, \cdots, \Theta_{n-1}$, conditional on $R = r$, we make use of the fact, proven in the appendix, that all orientations of the nearest neighbour are equally likely. Hence, if A is any subset of the sphere $\sum_1^n x_j^2 = r^2$, then

$$P\{n.n. \in A | R = r\} = \frac{\text{surface area (A)}}{S_n(r)} \quad .$$

So,

$$P\{\Theta_1 \leq \theta_1, \Theta_2 \leq \theta_2, \cdots, \Theta_{n-1} \leq \theta_{n-1} | R = r\}$$
$$= \frac{\text{surface area } \left\{(\Theta_1, \cdots, \Theta_{n-1}): \ -\frac{\pi}{2} \leq \Theta_j \leq \theta_j, \ j = 1, \cdots, n-2, \ 0 < \Theta_{n-1} \leq \theta_{n-1}\right\}}{S_n \cdot r^{n-1}} \quad .$$

Now the area of a region A of a sphere of radius r in n dimensions is

$$\int_{A'} dS = \int_{A'} \frac{r}{\sqrt{r^2 - x_1^2 - \cdots - x_{n-1}^2}} \, dx_1, \cdots, dx_{n-1} \qquad (3)$$

where A' is the projection of A onto the $n-1$ dimensional subspace $(X_1, \cdots, X_{n-1})$. Complications which may occur when the hyperplane $X_n = 0$ cuts the interior of the region A (i.e., we are integrating over parts of both the upper and lower half spheres) can be avoided by letting $X_n = 0$ partition the region into two subregions and integrating over each separately.

We will find it convenient to evaluate (3) in polar coordinates. Since the integration is over the $n-1$ dimensional subspace $(X_1, \cdots, X_{n-1})$, we introduce the variables of integration $\rho, \phi_2, \cdots, \phi_{n-1}$. ρ is the length of the projection of $(X_1, \cdots, X_n)$ onto the space $(X_1, \cdots, X_{n-1})$. Hence $\rho = R \cos \Theta_1$ (with some freedom of notation). The projection of $(X_1, \cdots, X_{n-1})$ onto $(X_1, \cdots, X_{n-2})$ space is exactly as previously described and so

$$\phi_2 = \Theta_2, \cdots, \phi_{n-1} = \Theta_{n-1} .$$

The transformation is given by

$$\begin{aligned} x_{n-1} &= \rho \sin \phi_2 \\ x_{n-2} &= \rho \cos \phi_2 \sin \phi_3 \\ &\vdots \\ x_1 &= \rho \cos \phi_2, \cdots, \cos \phi_{n-2} \sin \phi_{n-1} . \end{aligned}$$

The limits of integration are thus

$$\begin{gathered} 0 \le \rho \le r \cos \theta_1 \\ -\frac{\pi}{2} \le \phi_j \le \theta_j, \quad j = 2, \cdots, n-2 \\ 0 \le \phi_{n-1} \le \theta_{n-1} . \end{gathered}$$

and the Jacobian becomes $\rho^{n-2} \cos \phi_2^{n-3} \cos \phi_3^{n-4} \cdots \cos \phi_{n-2}$. So (3) becomes

$$\int_0^{r\cos\theta_1} \int_{-\frac{\pi}{2}}^{\theta_2} \cdots \int_{-\frac{\pi}{2}}^{\theta_{n-2}} \int_0^{\theta_{n-1}} \frac{r}{\sqrt{r^2-\rho^2}} \rho^{n-2} \cos\phi_2^{n-3} \cos\phi_3^{n-4} \cdots \cos\phi_{n-2} \, d\phi_{n-1} \, d\phi_{n-2} \cdots d\phi_2 \, d\rho$$

$$= r \int_0^{r\cos\theta_1} \frac{\rho^{n-2}}{\sqrt{r^2-\rho^2}} d\rho \prod_{j=2}^{n-2} \int_{-\frac{\pi}{2}}^{\theta_j} [\cos \phi_j]^{n-j-1} \int_0^{\theta_{n-1}} d\theta_{n-1}$$

$$= r^{n-1} \theta_{n-1} \prod_{j=1}^{n=2} \int_{-\frac{\pi}{2}}^{\theta_j} [\cos \phi_j]^{n-j-1} d\phi_j \quad .$$

Since the total surface area of sphere of radius r in n dimensions is $S_n r^{n-1}$ then

$$\text{Prob}\{\Theta_j \le \theta_j ; \; j = 1, \cdots, n-1\} = S_n^{-1}\theta_{n-1} \prod_{j=1}^{n-2} \int_{-\frac{\pi}{2}}^{\theta_j} [\cos \phi_j]^{n-j-1} d\phi_j$$

and the joint density of $\Theta_1, \cdots, \Theta_{n-1}$ conditional on R· is

$$S_n^{-1} \prod_{j=1}^{n-2} [\cos \theta_j]^{n-j-1} = f_{\Theta_1, \cdots, \Theta_{n-1} | R}(\theta_1, \cdots, \theta_{n-1} | r) \quad .$$

Finally, the joint density of R, $\Theta_1, \cdots, \Theta_{n-1}$ is

$$f(r, \theta_1, \cdots, \theta_{n-1}) = \lambda \cdot V_n \cdot S_n^{-1} \cdot n \cdot r^{n-1} e^{-\lambda \cdot V_n r^n} \cdot \prod_{j=2}^{n-2} (\cos \theta_j)^{n-j-1}$$

for $r \ge 0$, $-\frac{\pi}{2} \le \theta_j \le \frac{\pi}{2}$ for $j = 1, \cdots, n-2$ and $0 \le \theta_{n-1} < 2\pi$. Now, transform from polar coordinates $(R, \Theta_1, \cdots, \Theta_{n-1})$ to Cartesian coordinates $(X_1, \cdots, X_n)$. The Jacobian is

$$\left\{ r^{n-1} \prod_{j=1}^{n-2} [\cos \theta_j]^{n-j-1} \right\}^{-1} \quad .$$

Thus the joint density of $X_1, \cdots, X_n$ is

$$\begin{aligned} f(x_1, \cdots, x_n) &= \lambda \cdot V_n S_n^{-1} \cdot n \cdot e^{-\lambda \cdot V_n \cdot (x_1^2 + \cdots + x_n^2)^{n/2}} \\ &= \lambda e^{-\lambda \cdot V_n \cdot (x_1^2 + \cdots + x_n^2)^{n/2}} \end{aligned}$$

since $S_n = n \cdot V_n$.

From the density we may make some observations. The coordinates of the nearest neighbour are not independently distributed except in the case $n = 2$. In that case, X_1 and X_2 each have a normal distribution with mean 0 and variance $1/2\pi\lambda$. However, they are uncorrelated for all n .

$$E(X_1 X_2) = \lambda \int_{R^n} x_1 x_2 e^{-\lambda \cdot V_n \cdot (x_1^2 + \cdots + x_n^2)^{n/2}} dx_1, \cdots, dx_n \quad .$$

Transforming to polar coordinates as before, we have

$$E(X_1 X_2) = \lambda \int_{R^n} r^2 \cos^2\theta_1 \cdots \cos^2\theta_{n-2} \cos\theta_{n-1} \sin\theta_{n-1} e^{-\lambda \cdot V_n \cdot r^n} \cdot r^{n-1} \cos\theta_1^{n-2} \cdots \cos\theta_{n-2} dr \cdot d\theta_1 \cdots d\theta_{n-2} d\theta_{n-1} \quad .$$

The integrals factor and in particular

$$\int_0^{2\pi} \cos\theta_{n-1} \sin\theta_{n-1} d\theta_{n-1} = 0 \quad .$$

Since the labelling of the coordinate axes was arbitrary we have that

$$E(X_i X_j) = 0, \; i \neq j \quad .$$

We can also derive an expression for the variance of X_j . First we note the following:

$$\begin{aligned} 1 &= \lambda \int_{R^n} e^{-\lambda \cdot V_n r^n} r^{n-1} (\cos \theta_1)^{n-2} \cdots \cos\theta_{n-2} dr \cdot d\theta_1 \cdots d\theta_{n-1} \\ &= \lambda \int_0^\infty r^{n-1} e^{-\lambda \cdot V_n r^n} dr \cdot \prod_{j=1}^{n-2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} [\cos \theta_j]^{n-j-1} d\theta_j \cdot \int_0^{2\pi} d\theta_{n-1} \\ &= \lambda \left(\frac{1}{\lambda \cdot n V_n} \right) \prod_{j=1}^{n-2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} [\cos \theta_j]^{n-j-1} d\theta_j \cdot (2\pi) \quad . \end{aligned}$$

So,

$$\prod_{j=1}^{n-2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (\cos \theta_j)^{n-j-1} d\theta_j = \frac{1}{2\pi} \cdot n \cdot V_n \quad . \tag{4}$$

Now, the variance of X_n is

$$E(X_n^2) = \lambda \int_{R^n} x_n^2 \, e^{-\lambda \cdot V_n \cdot (x_1^2 + \cdots + x_n^2)^{n/2}} \, dx_1, \cdots, dx_n \quad .$$

Changing to polar coordinates this becomes

$$\lambda \int_0^\infty r^{n+1} e^{-\lambda \cdot V_n r^n} dr \cdot \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^2\theta_1 \, \cos^{n-2}\theta_1 \, d\theta_1 \cdot \prod_{j=2}^{n-2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (\cos \theta_j)^{n-j-1} d\theta_j \int_0^{2\pi} d\theta_{n-1} \quad . \tag{5}$$

But

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^2\theta_1 \, \cos^{n-2}\theta_1 \, d\theta_1 = \frac{1}{n} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^{n-2}\theta_1 \, d\theta_1 \, , \tag{6}$$

so using (4) and (6), (5) becomes

$$E(X_n^2) = \frac{1}{n} \left(\frac{1}{\lambda \cdot V_n}\right)^{2/n} \Gamma\left(\frac{2}{n}+1\right) \tag{7}$$

where $\Gamma(\cdot)$ is the gamma function. By symmetry, the same result holds for $E(X_j^2)$, $j = 1, \cdots, n$. In fact, as a check we note that $E(X_j^2) = \frac{1}{n} E\left(\sum_1^n X_i^2\right) = \frac{1}{n} E(R^2)$ $= \frac{1}{n} \left(\frac{1}{\lambda \cdot V_n}\right)^{2/n} \Gamma\left(\frac{2}{n}+1\right)$.

Another observation of potential statistical utility is that although $X_1, X_2, \cdots, X_n$ are not iid $\mathbb{N}(0,\sigma^2)$ (except in the case $n=2$), the r.v.'s $\Theta_1, \Theta_2, \cdots, \Theta_{n-1}$ have the same distribution in our setting as in the case when $X_1, \cdots, X_n$ are iid $\mathbb{N}(0,\sigma^2)$. Further, $\Theta_1, \cdots, \Theta_{n-1}$ are statistically independent of $R = \sqrt{\Sigma X_i^2}$ in each case. It follows, since the statistic $F = n\bar{x}^2/s^2$ (and hence $t = \sqrt{n}\,\bar{x}/s$) is a function only of $\Theta_1, \cdots, \Theta_{n-1}$, that it has the same distribution in both settings ($(n-1)s^2 = r^2 - n\bar{x}^2$) . Hence the usual tables could be used in our Poisson setting.

4. Extension to the k-th Nearest Neighbour

To determine the joint density of the coordinates of the k-th nearest neighbour we will, as before, determine the density of its polar coordinates and transform to the Cartesian coordinates.

To determine the density of distance to the k-th nearest neighbour, we first construct the joint density of the distance to the nearest, R_1, second nearest, R_2, $\cdots$, and k-th nearest neighbours, R_k .

$$P\{R_j > r_j \mid R_1 = r_1, \cdots, R_{j-1} = r_{j-1}\} = P\{\text{no individuals between the sphere } \Sigma x_j^2 = r_{j-1}^2 \text{ and } \Sigma x_j^2 = r_j^2\}$$

$$= e^{-\lambda V_n (r_j^n - r_{j-1}^n)} \quad .$$

Hence the joint density of R_1, R_2, $\cdots$, R_k is

$$f_{R_1, R_2, \cdots, R_k}(r_1, r_2, \cdots, r_k) = (n \cdot \lambda \cdot V_n)^k r_1^{n-1} \cdots r_k^{n-1} e^{-\lambda \cdot V_n \cdot r_k^n} ,$$

$$0 \le r_1 \le r_2 \le \cdots \le r_k < \infty \quad .$$

Then by integrating out r_1, r_2, $\cdots$, r_{k-1} we obtain

$$f_{R_k}(r_k) = \frac{n(\lambda \cdot V_n)^k}{(k-1)!} r_k^{nk-1} e^{-\lambda \cdot V_n \cdot r_k^n} , \quad 0 < r_k < \infty \quad .$$

The previous arguments used to obtain the joint density of Θ_1, $\cdots$, Θ_{n-1} conditional on R remain unchanged when we condition on R_k and hence the joint density of R_k, Θ_1, $\cdots$, Θ_{n-1} is

$$\frac{n(\lambda \cdot V_n)^k}{(k-1)!} r_k^{n \cdot k-1} e^{-\lambda \cdot V_n \cdot r_k^n} S_k^{-1} \prod_{j=2}^{n-2} [\cos \theta_j]^{n-k-1} \quad .$$

Finally, transforming to Cartesian coordinates yields the joint density of the coordinates $(X_1, \cdots, X_n)$ of the k-th nearest neighbour.

$$\frac{\lambda^k V_n^{k-1}}{(k-1)!} (x_1^2+\cdots+x_n^2)^{\frac{n(k-1)}{2}} e^{-\lambda\cdot V_n\cdot(x_1^2+\cdots+x_n^2)^{n/2}} ,$$
$$-\infty < x_j < \infty \quad (j=1,\cdots,n) .$$

References

[1] Kendall, M. G. (1961). The Geometry of n-Dimensions. Charles Griffin, London.

6. Appendix

We will prove the lemma used in Section 3.

Lemma. Conditional on the event that the nearest neighbour (n.n.) is r units from the origin, the density of its position is uniform on the sphere $\Sigma x_j^2 = r^2$.

Proof: Define the following regions in R^n . Let L be the region contained within the sphere $\Sigma x_j^2 = r^2$ and M be the region bounded by the sphere $\Sigma x_j^2 = r^2$ and $\Sigma x_j^2 = (r+\Delta r)^2$. Let N be a subset of M as follows. Take any region A on the sphere $\Sigma x_j^2 = r^2$. Connect each point on the boundary of A to the origin with a straight line. Then let N be the intersection with M of the half-cone formed in this way.

The volume of L, denoted by V(L), is $V_n\cdot r^n$. Similarly, $V(M) = V_n\cdot[(r+\Delta r)^r - r^n]$ denotes the volume of M and $V(N) = S(A)/S_n(r)\cdot V(M)$ is the volume of N where S(A) denotes the surface area of A and $S_n(r)$ the surface area of a sphere of radius r . By definition,

$$P\{n.n.\epsilon A | R=r\} = \lim_{sr\to 0} \frac{P\{n.n.\epsilon N\}}{P\{n.n.\epsilon M\}} \tag{8}$$

if this limit exists. By properties 1) and ii) of the Poisson process

$$P\{n.n.\in M\} = P\{\text{no elements in } L\} \cdot P\{\text{at least one in } M\}$$

$$= e^{-\lambda \cdot V(L)} \left(1 - e^{-\lambda \cdot V(M)}\right) . \tag{9}$$

To compute the numerator of (8) we note that

$$P\{n.n.\in N\} = P\{\text{no elements in } L\} \cdot P\{1 \text{ element in } N\} \cdot P\{\text{no elements in } M-N\}$$
$$+ P\{\text{no elements in } L\} \cdot P\{2 \text{ or more elements in } M \text{ with } n.n.\in N\} . \tag{10}$$

The second term of this expression, when divided by (9), becomes negligible as $\Delta r \to 0$ since

$$P\{\geq 2 \text{ elements in } M \text{ and } n.n.\in N\} \cdot P\{\text{no elements in } L\}$$
$$\leq P\{\geq 2 \text{ elements in } M\} \cdot P\{\text{no elements in } L\}$$
$$= \left\{1 - e^{-\lambda \cdot V(M)}\left(1 + \lambda \cdot V(M)\right)\right\} \cdot \left\{e^{-\lambda \cdot V(L)}\right\} . \tag{11}$$

With an application of l'Hôpital's rule, $\lim\limits_{\Delta r \to 0}$ of (11) divided by (9) is 0 .

Returning to (10), the first term is

$$P\{\text{no elements in } L\} \cdot P\{1 \text{ element in } N\} \cdot P\{\text{no elements in } M-N\}$$

$$= e^{-\lambda \cdot V(L)} \cdot \frac{S(A)}{S_n(r+\Delta r)} \lambda \cdot V(M) \exp\left\{- \frac{S(A)}{S_n(r+\Delta r)} \cdot \lambda \cdot V(M)\right\}$$
$$\cdot \exp\left\{- \frac{S_n(r+\Delta) - S(A)}{S(A)} \cdot \lambda \cdot V(M)\right\}$$

$$= e^{-\lambda \cdot V(L)} \cdot \frac{S(A)}{S_n(r+\Delta r)} \cdot \lambda \cdot V(M)\; e^{-\lambda \cdot V(M)} . \tag{12}$$

Dividing (12) by (9), letting $\Delta r \to 0$ and using l'Hôpital's rule, we obtain

$$P\{n.n. \in A | R = r\} = \frac{S(A)}{S_n(r)} = \frac{S(A)}{V_n \cdot r^n} \quad .$$

Hence, the conditional density on the sphere must be uniform.